AUTOMATIC TIME-ACTIVITY CLASSIFICATION BASED ON GLOBAL POSITIONING DATA

Jun Wu, University of California, Irvine, USA
Chengsheng Jiang, University of California, Irvine, USA
Douglas Houston, University of California, Irvine, USA
Ralph Delfino, University of California, Irvine, USA
Dean Baker, University of California, Irvine, USA

Background and Aims: Air pollution epidemiological studies are increasingly using global positioning system (GPS) to collect time-location data. However, remarkable uncertainties in GPS data create challenges for reliably characterizing time-activity patterns from the raw GPS data. We aimed to develop models to classify people's major time-activity patterns directly from the GPS tracking.

Methods: We collected 131 person-days of GPS time-activity data in 2008 for 47 residents in Southern California. Time-activity patterns were coded for each subject based on GPS recordings, time-activity diary logs, and follow-up interviews. Using similar methods, 7-day validation data were collected from three research staff in 2010. We developed, evaluated, and compared two models for extracting time-activity patterns (i.e. indoor static, outdoor static, outdoor walking, and in-vehicle travel) from the GPS data. A rule-based model developed user-defined rules based on time, speed, and spatial relationship of the GPS points, while a logistic model trees (LMT) model required minimal user-input by using existing software that combined a tree structure with logistic regression.

Results: Indoor static, outdoor static, outdoor walking and in-vehicle travel accounted for 82.7%, 6.1%, 3.2% and 7.2% of the points in the main dataset, respectively. Both the rule-based and the LMT models correctly classified more than 90% of reported indoor static points and more than 70% of reported in-vehicle travel points, with Type 1 error (over-estimation) and Type 2 error (underestimation) <9% for indoor static classification and <29% for in-vehicle travel classification. Both models performed the worst in identifying outdoor static points. The performance of the LMT model improved significantly when using training data from the 7-day validation dataset likely because of more accurate diary data.

Conclusion: Our models can successfully identify static and in-vehicle travel points from raw GPS data. It is more challenging to separate outdoor static from indoor static points.